Overview of CCS-5

Simulation Based Modeling for Threat Reduction

Nick Hengartner

Stephan Eidenbenz, Anders Hansson, Gabriel Istrate
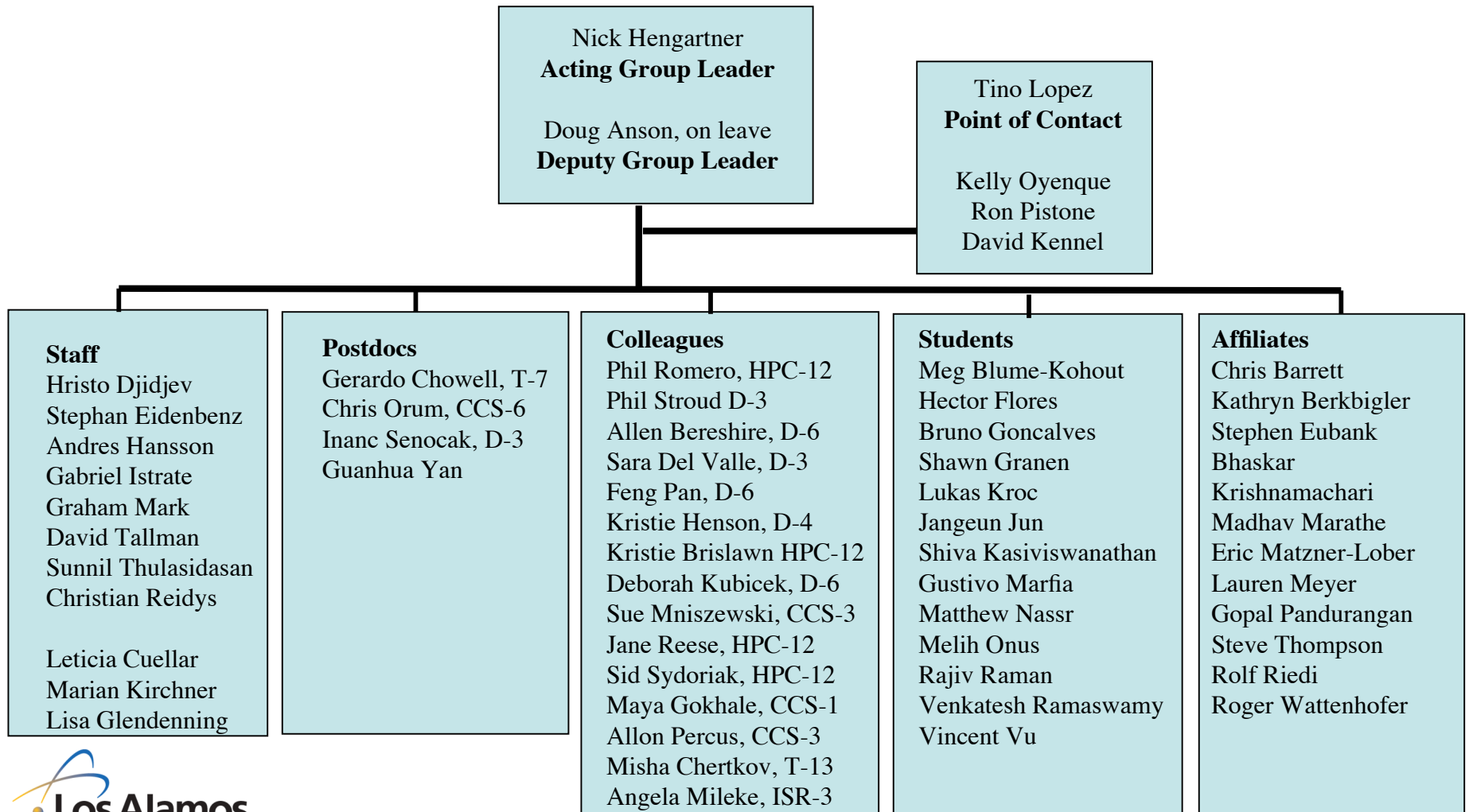
Mission and Vision

Mission
Develop and apply unique
technical and programmatic capabilities
in simulation-based modeling to support
science based predictions, information
integration, and analysis for threat
reduction.

Vision
Be the lead applied math and computer
science group at LANL for scientific and
engineering foundations of large scale agent
based simulation modeling to support
science based predictions and information
integration.

Los Alamos
NATIONAL LABORATORY
EST.1943

People

Nick Hengartner
Acting Group Leader

Doug Anson, on leave
Deputy Group Leader

Tino Lopez
Point of Contact

Kelly Oyenque
Ron Pistone
David Kennel

Staff
Hristo Djidjev
Stephan Eidenbenz
Andres Hansson
Gabriel Istrate
Graham Mark
David Tallman
Sunnil Thulasidasan
Christian Reidys

Leticia Cuellar
Marian Kirchner
Lisa Glendenning

Postdocs
Gerardo Chowell, T-7
Chris Orum, CCS-6
Inanc Senocak, D-3
Guanhua Yan

Colleagues
Phil Romero, HPC-12
Phil Stroud D-3
Allen Bereshire, D-6
Sara Del Valle, D-3
Feng Pan, D-6
Kristie Henson, D-4
Kristie Brislawn HPC-12
Deborah Kubicek, D-6
Sue Mniszewski, CCS-3
Jane Reese, HPC-12
Sid Sydoriak, HPC-12
Maya Gokhale, CCS-1
Allon Percus, CCS-3
Misha Chertkov, T-13
Angela Mileke, ISR-3

Students
Meg Blume-Kohout
Hector Flores
Bruno Goncalves
Shawn Granen
Lukas Kroc
Jangeun Jun
Shiva Kasiviswanathan
Gustivo Marfia
Matthew Nassr
Melih Onus
Rajiv Raman
Venkatesh Ramaswamy
Vincent Vu

Affiliates
Chris Barrett
Kathryn Berkbigler
Stephen Eubank
Bhaskar
Krishnamachari
Madhav Marathe
Eric Matzner-Lober
Lauren Meyer
Gopal Pandurangan
Steve Thompson
Rolf Riedi
Roger Wattenhofer

Los Alamos
NATIONAL LABORATORY
— EST.1943 —

Scientific Productivity (2005-2006)

40+ peer reviewed/invited scientific papers/books (research by the pound)

1. Preventing bandwidth abuse at the router
2. Evaluating the effect of interventions against pandemic influenza.
3. Evaluating experiments for estimating the bit failure cross-section of semiconductors using a colored spectrum neutron beam.
4. Transmission Dynamics of the Great Influenza Pandemic of 1918 in Geneva, Switzerland: Assessing the Effects of Hypothetical Interventions.
5. Comparing deployments of moving and fixed air samplers for detecting and mapping biological releases.
6. Critical response to post-outbreak vaccination against foot-and-mouth disease.
7. Using the LANSCE Irradiation Facility to Predict the Number of Fatal Soft Errors in one of the World's Fastest Supercomputer.
8. Estimation of the reproductive number of Spanish fly epidemic in Geneva, Switzerland.
9. Variable selection for Gaussian Process Models in Computer Experiments.
10. Efficient strategy of low statistics nuclear searches.
11. Information Extraction for Muon Radiography.
12. Measures of Inequitable Proximity to Major Highways in NYC.
13. Adaptive estimation for inverse problems with noisy Operators
14. Structural Learning with time-varying components: tracking the cross section of financial time series.
15. Rate Optimal Estimation with the Integration Method in the Presence of Many Covariates
16. Predicting the Number of Fatal Soft Errors in Los Alamos National Laboratoty's ASC Q Supercomputer.
17. Demographic convergence in epidemics on realistic urban social networks.
18. Counting Preimages of TCP Reordering Patterns.
19. Combinatorics of TCP Reordering
20. A Hybrid CPU/FPGA Implementation of the TRANSIMS Micro-simulator
21. A Methodology for Semantic Compression of TCP Traces
22. Strong Edge Coloring for Channel Assignment in Wireless Radio Networks
23. Bluetooth Worms: Models Dynamics, and Defense Implications
24. Finding Minimum Hidden Guard Sets in Polygons—Tight Approximability Results
25. MIITS: Multi-scale Integrated Information and Telecommunication System
26. A Linear-Time Algorithm for Finding a Maximal Planar Subgraph
27. Planar Crossing Numbers of Genus g Graphs
28. Approximate Shortest Path Queries on Weighted Polyhedral Surfaces
29. A Discrete Dynamical Systems Framework for Packet-Flow on Networks
30. On a New Class of Load Balancing Network Protocols
31. Low-Complexity Iterative Detection Based on Limited Bi-Directional Trellis Search
32. On Asynchronous Cellular Automata
33. Generalized MSP Detection of Continuous Phase Modulation over Unknown ISI Channels
34. When Statistical Mechanics Meets Computation

Program committee:
IEEE MASS, SPTA, SecPerU, ACM SIGMOBILE, LACSI

Editorship:
Journal Parallel Processing Letters

Adjunct academic positions:
Carlton U., Simon Fraser U.

Research in support of Threat Reduction

Infrastructure Risk Assessment

Transims, Transportation Modeling, Evacuation, Emergency Response UIS, Urban Infrastructure Modeling
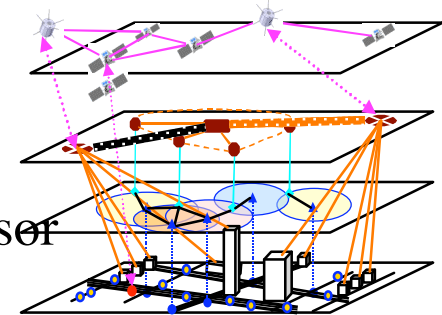
Public Health Modeling

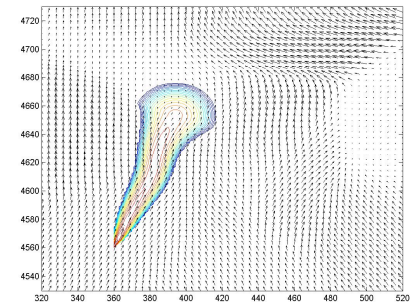Epidemic modeling on Social Networks, Computer worms propagation modeling,

Detection, Sensing and Communication

MIITS, Communication Modeling, Protocol design game theory, Wireless Mobile Sensor Networks, Sensor Deployment design, Model Driven Measurement, Muon tomography

General methods

Generic cities, Unifying Framework for UIS, Episims and MIITS.

Capabilities

Basic Science   Algorithmics, Complexity Theory, Graph Theory, Computation Geometry, Algebra, Combinatorics, Physics of Algorithm, Computer Networking, Game Theory

Simulation Science   HPC-based simulations: Discrete and Agent Based Simulations of socio-technical systems, Sequential Dynamic Modeling, Simulation Engine

Information Science   Network Analysis, Stochastic Modeling, Science Based Predictions, Sensor Networks, Signal processing, formal methods.
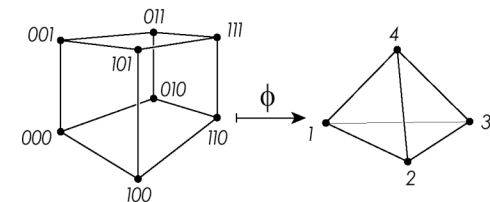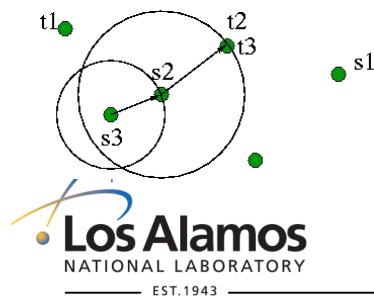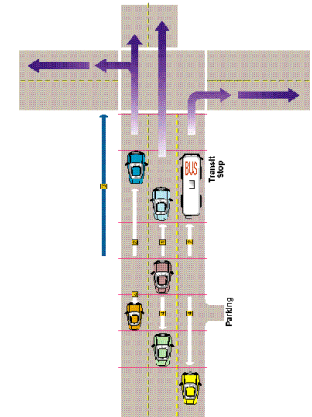
Funding
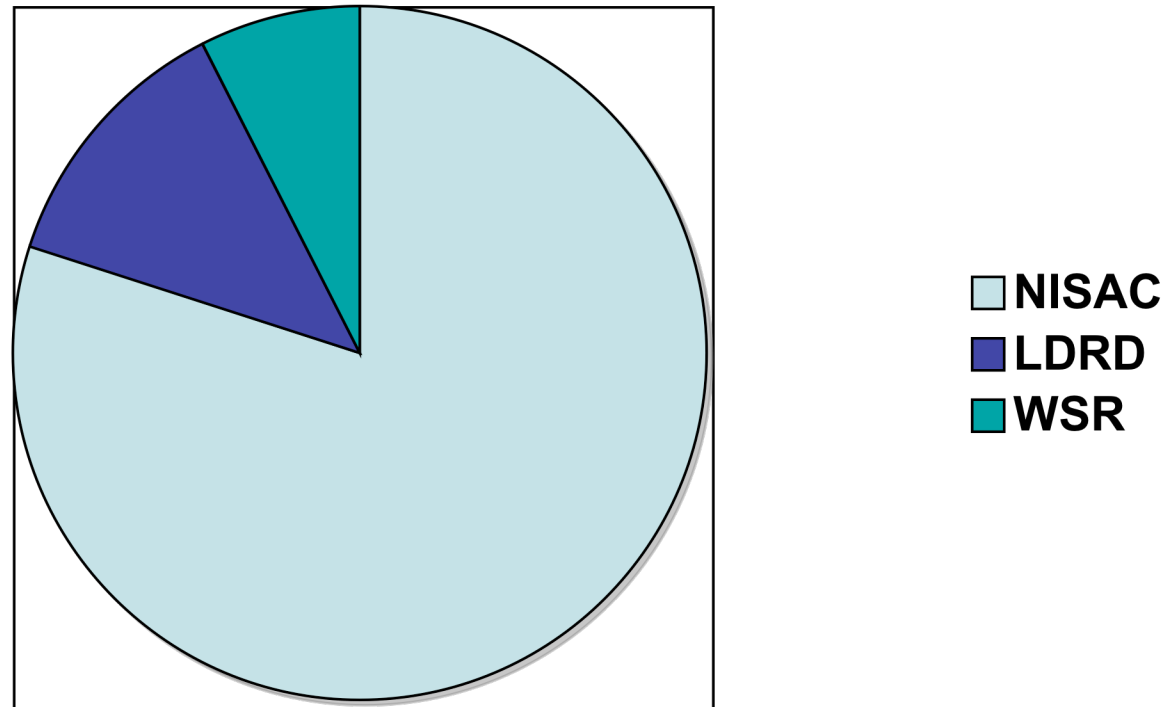
Philosophy of Socio-technical simulations

- Simulation **does not** predict the exact course of reality

- Simulation provides
  - ♣ a few likely **realizations** of a stochastic process in an enormous space
  - ♣ capability to model **interactions** at a **disaggregated** level

- Outcomes based on **aggregate** features of results are
      predictable
- Compare to Statistical Mechanics: can molecular
      dynamics simulations recover
  - ♣ trajectories of individual molecules?
  - ♣ heat capacity?

Entity → State → State Change
Interaction → Composition of Interactions → System Realizations

SDS
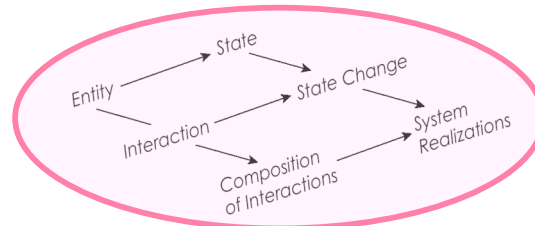
Simulation as Data integration tool

- Generic simulation-based methods for **integrating data**
  - ♣ diverse commercial and public data sources, expert knowledge

    (can work in data poor environments)
  - ♣ Survey data, Census data, IP Databases, Road networks, etc.

- Generic lightweight parametric methods for **agent representation**

- Generic methods for **storing and regenerating** large

    real as well as synthetic data
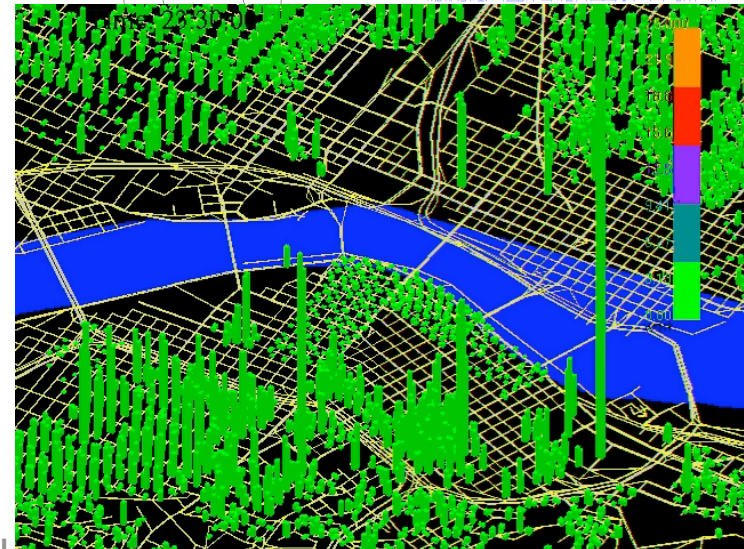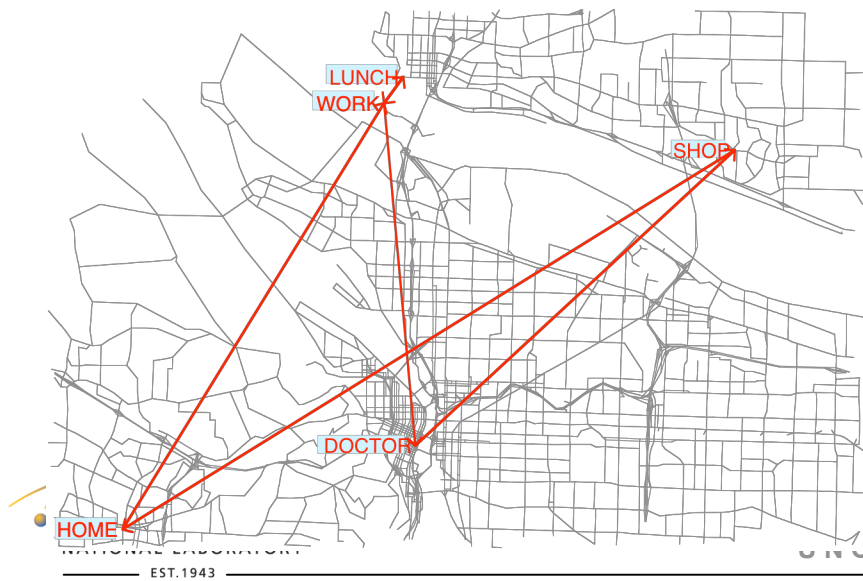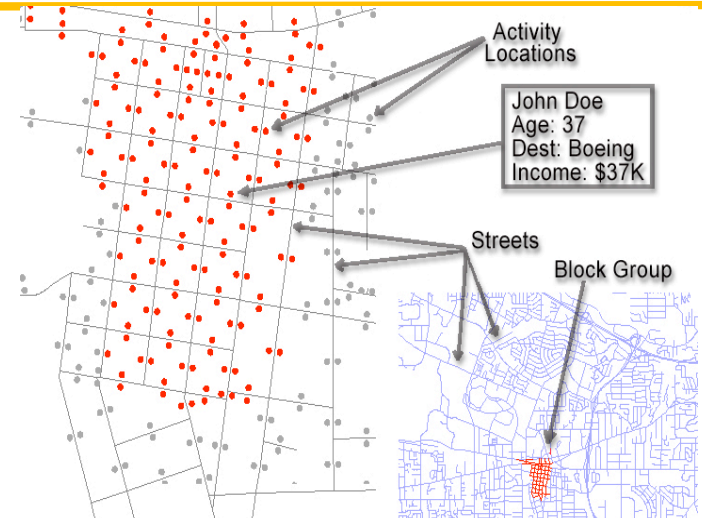 - ♣ Used in reconstructing important dynamics (reduced, ultrafast simulations)

Los Alamos
NATIONAL LABORATORY
EST.1943

NNSA

Urban Infrastructure Suite (UIS)
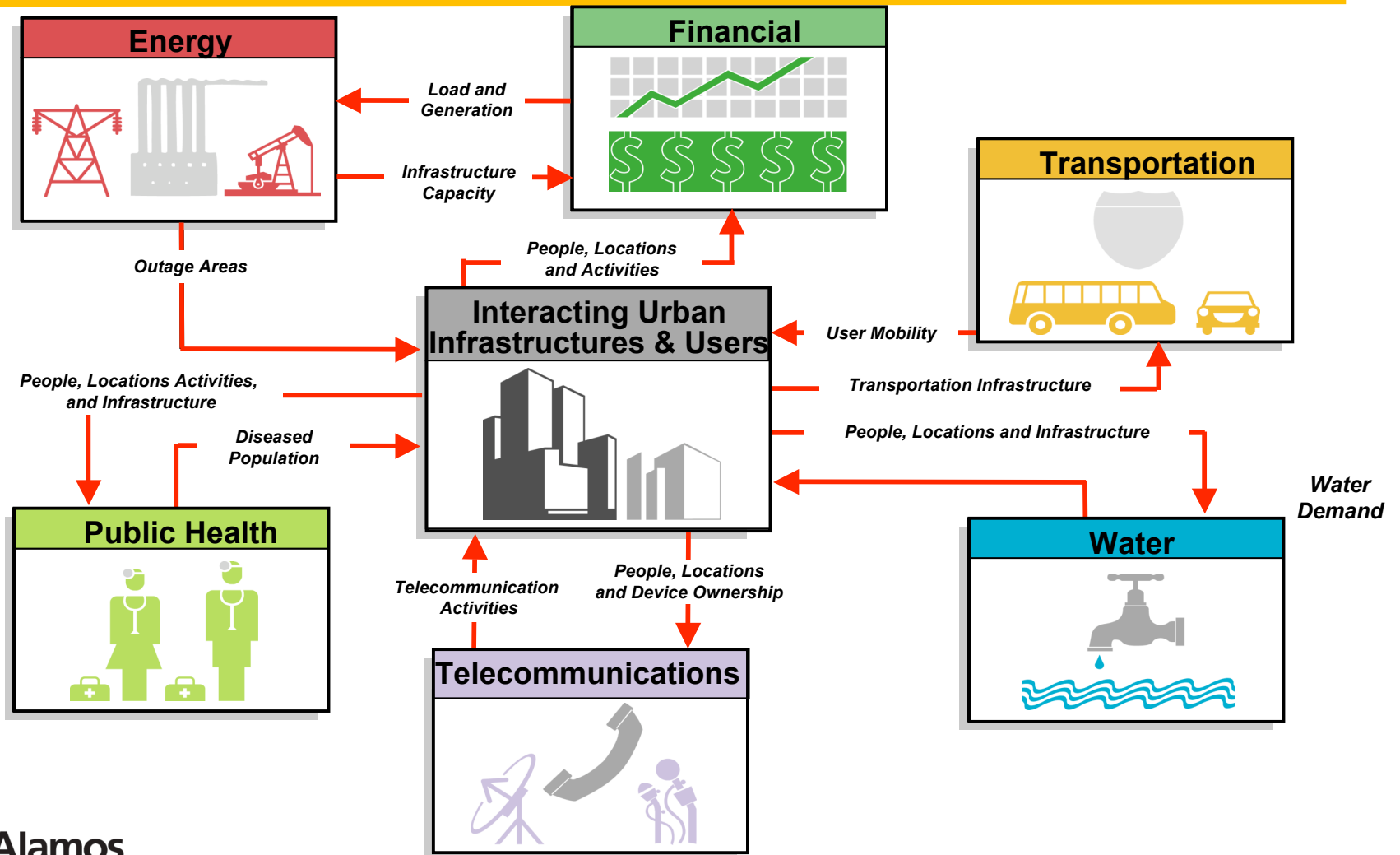
Synthetic Representation of Mobile Urban Population

Age	26	26	7
Income	$27k	$16k	$0
Status	worker	worker	student
Automobile			

Activity Locations

John Doe
Age: 37
Dest: Boeing
Income: $37K

Streets

Block Group

LUNCH
WORK
SHOP
DOCTOR
HOME

NATIONAL LABORATORY
EST. 1943

NNSA
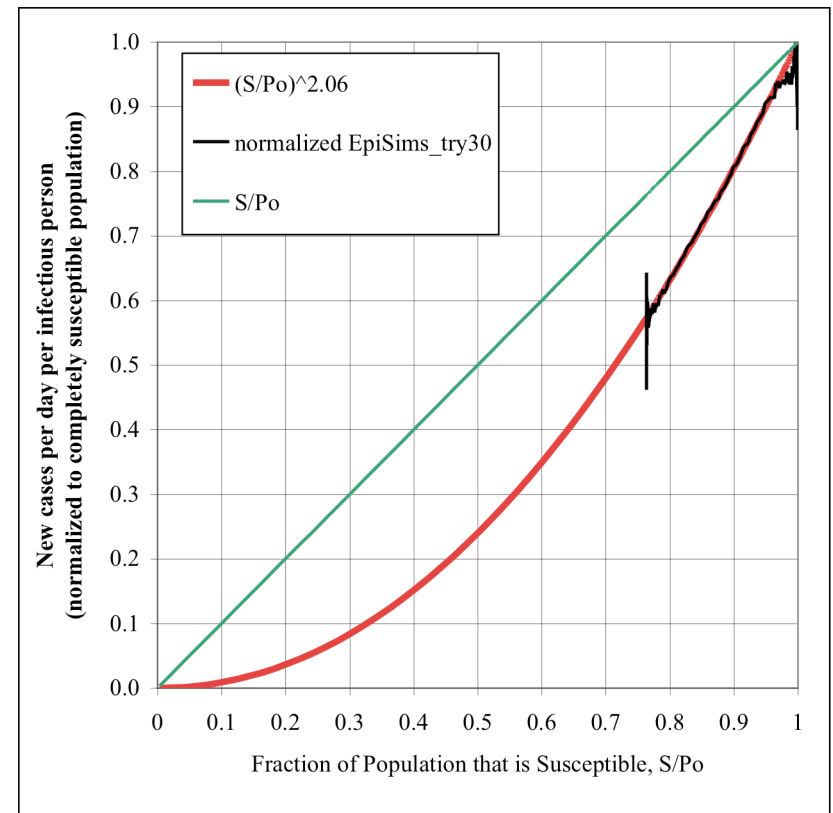
New Domain Science: Power Law Mixing

- Early infections occur disproportionably higher in persons with larger number of contacts

- As the epidemic progresses, the remaining susceptible & infectious people have fewer contacts

- New cases per day

becomes $n = I(R_0 / \tau_I)(S / P_0)$

$$n_{LA} = I(R_0 / \tau_I)(S_0 / P_0)(S / S_0)^{2.1}$$

Los Alamos
NATIONAL LABORATORY
EST.1943

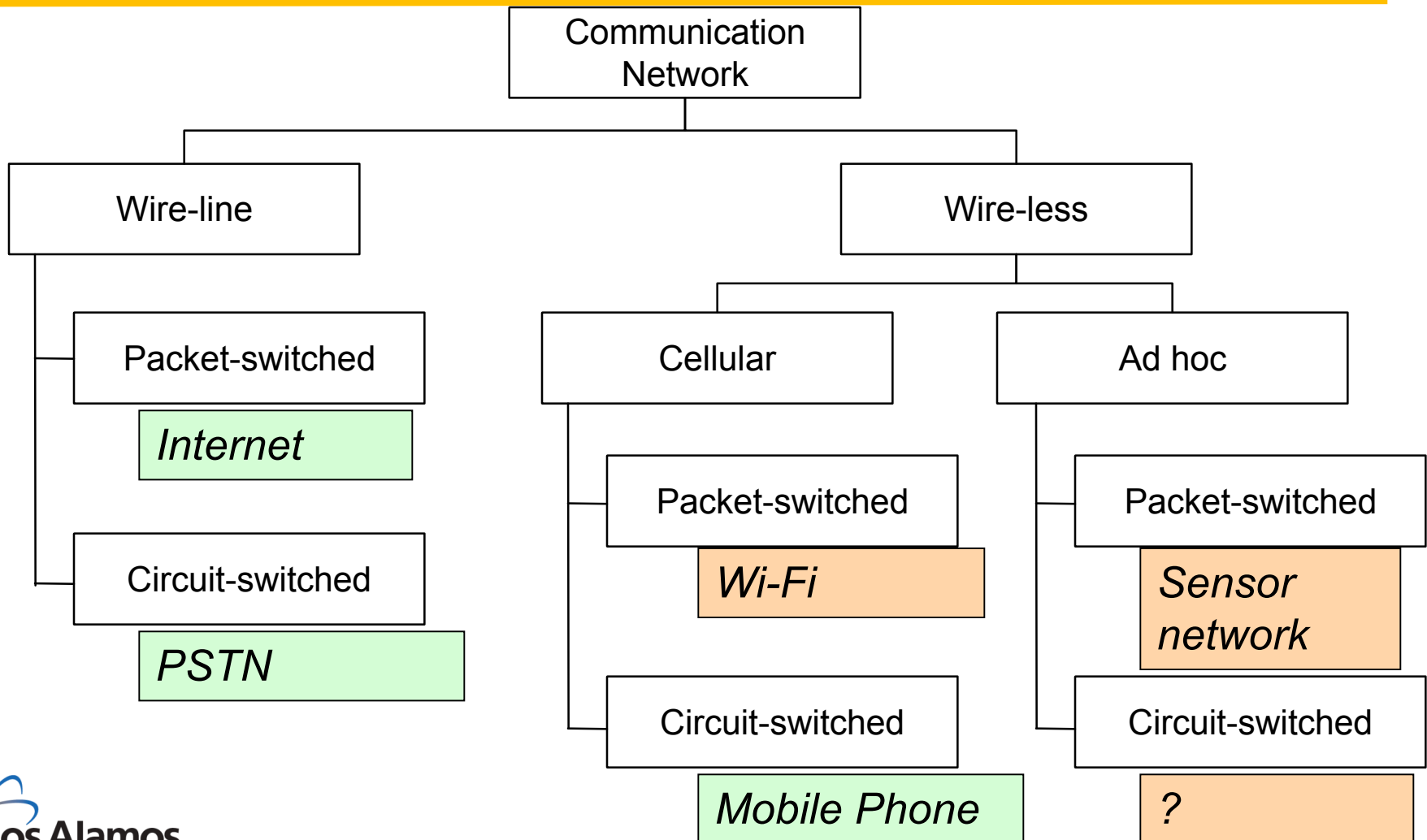# Multi-scale Integrated Information and Telecommunications System (MIITS)

Stephan Eidenbenz, CCS-5

L. Cuellar
V. Ramaswamy
J. Jun
L. Glendenning
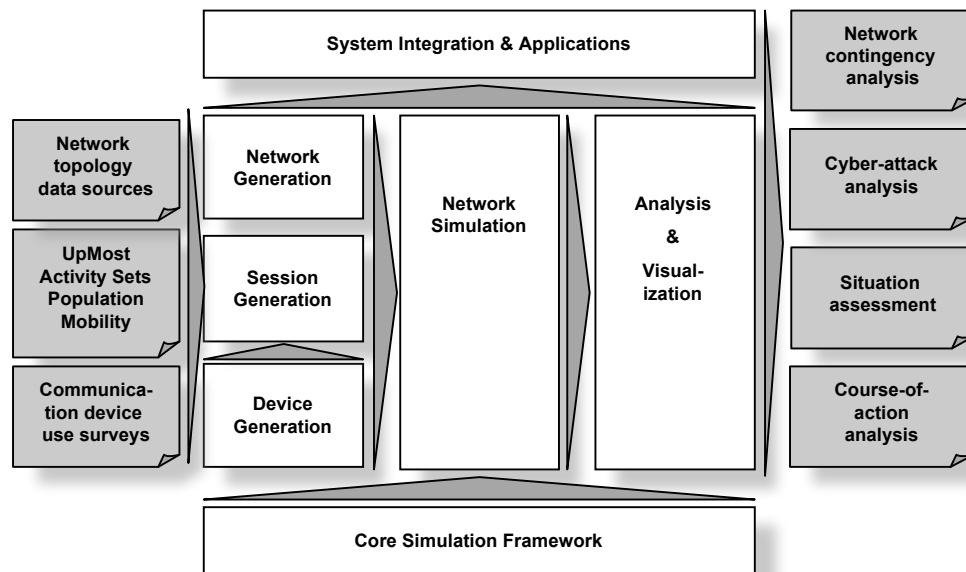C. Reidys
P. Datta
F. Pan

S. Thulasidasan
G. Marfia
H. Flores
G. Mark
N. Hengartner
A. Berscheid
P. Romero

G. Yan
M. Nassr
L. Kroc
C. Tallman
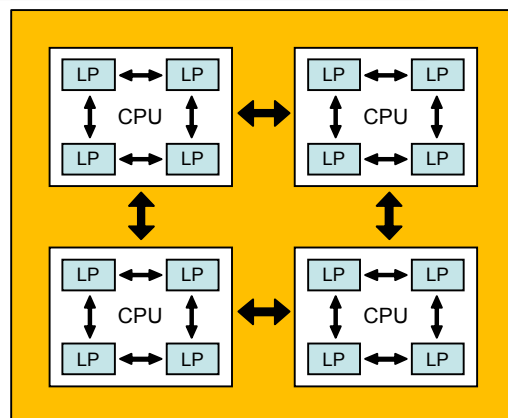J.P. Smith
R. Waupotitsch
a.o.

Communication Network Types

Communication Network
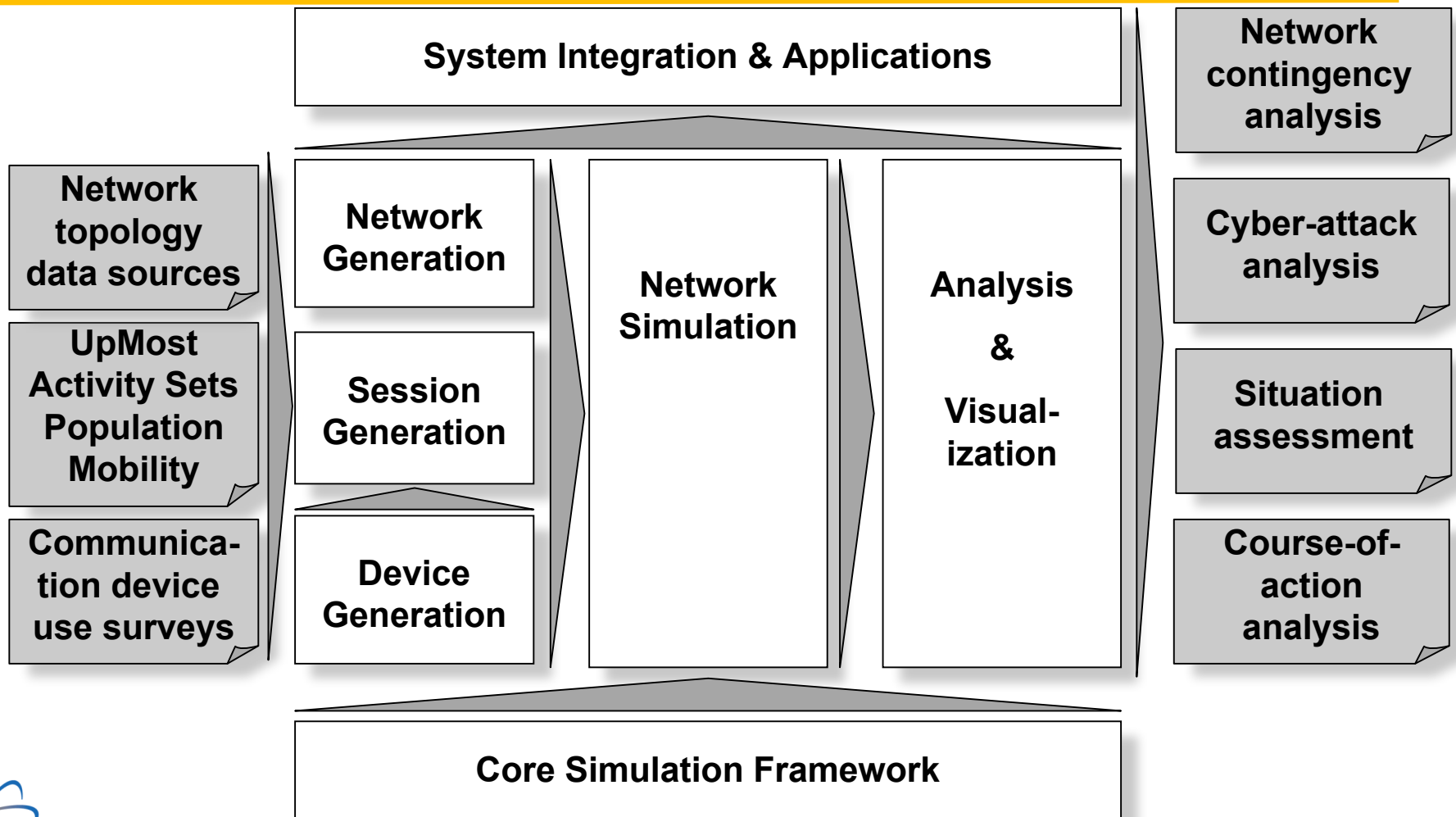
Wire-line

Wire-less

Packet-switched

Internet

Circuit-switched

PSTN

Cellular

Ad hoc

Packet-switched

Wi-Fi

Packet-switched

Sensor network

Circuit-switched

Mobile Phone

Circuit-switched

?

Los Alamos
NATIONAL LABORATORY
— EST. 1943 —

NNSA

MIITS Design Principles

System Integration & Applications

Network topology data sources

UpMost Activity Sets Population Mobility

Communication device use surveys

Network Generation

Session Generation

Device Generation

Network Simulation

Analysis & Visual-ization

Network contingency analysis

Cyber-attack analysis

Situation assessment

Course-of-action analysis

Core Simulation Framework

- MIITS modules Session Generation, Network Simulation, and Analysis designed for use on Supercomputer

- Realistic input creation modules tied to UpMoST and TRANSIMS

- Distributed discrete event-driven simulation technology ensures efficient scaling on any architecture

- Communication network simulation-specific process synchronization and dynamic load-dependent partitioning schemes and memory footprint reduction techniques currently in development and implementation

- Competitive simulators not designed to scale to this level

MIITS Architecture Overview: Input/Scalability

System Integration & Applications

Network contingency analysis

Network topology data sources

UpMost Activity Sets Population Mobility

Communica-tion device use surveys

Network Generation

Session Generation

Device Generation

Network Simulation

Analysis & Visual-ization

Cyber-attack analysis

Situation assessment

Course-of-action analysis

Core Simulation Framework

MIITS Scientific Foundation

- More than 30 MIITS-related articles published in peer-reviewed venues: e.g. ACM MobiCom, ACM/SIAM SODA, ACM Mobihoc, MONET, IEEE PerCom (acceptance rates vary between 8% and 35%)

- Current pipeline: 10+ publications

- Team members serve on TPC: e.g. IEEE Infocom 2007, IEEE MASS 2006, ACM Mobihoc 2006, SecPerU 2006, Mobihoc 2005, DIAL-M-POMC 2005, SecPerU 2005, DIAL-M 2003

- Team members review for top journals on regular basis: e.g. ToN, TMC, JPDS, JSAC, ADHOC, MONET

MIITS Selected Recent Results

- ACM Mobicom 2006: OURS- Optimal Unicast Routing Systems in Non-Cooperative Wireless Networks
  - Game-theoretic mechanisms for routing in ad hoc networks with provably minimum overpayment ratios for dominant-strategy and Nash equilibria

- IEEE/ACM MASCOTS 2006: Sluggish Calendar Queues
  - Novel event-queue data structure optimized for discrete simulation of communication networks by exploiting statistics on event streams

- ACM MobiHoc 2006: Algorithmic Aspects of Communication in Ad-Hoc Networks with Smart Antennas
  - Combinatorial characterization and approximation algorithms for interference minimization in smart antenna ad hoc networks

- Mobile Networks & Applications (Springer) 2005: Parametric Probabilistic Sensor Network Routing in Sensor Networks
  - Multi-path limited-flooding, tiny-state routing protocol for sensor networks; currently being tested on real motes

- IEEE PerCom Workshops 2005: Maneuverable Relays to Improve Energy Efficiency in Sensor Networks
  - Hierarchical routing approach for sensor networks using mobile relays

Parametric Probabilistic Sensor Network Routing Protocols

[C. Barrett, S. Eidenbenz, L. Kroc, M. Marathe, J.P. Smith, WSNA 2003, MONET 2005, ACM SAC 2005]

Family of routing protocols characterized by retransmission probability function

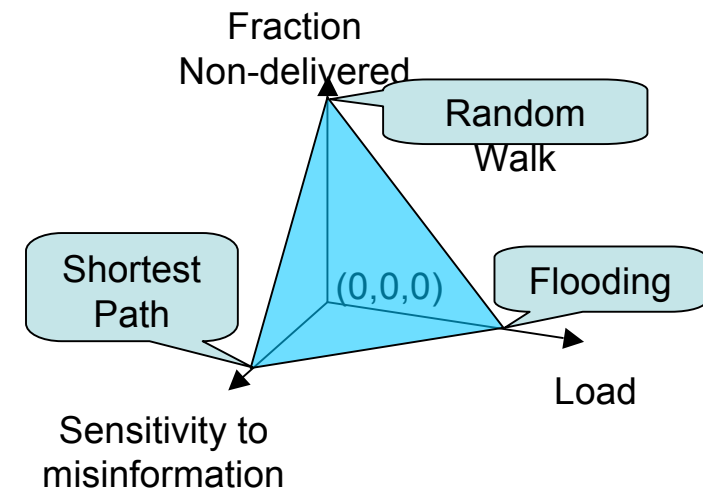Performance tested in simulation and implemented on 49 sensor motes test-bed

Example: Directed Transmission

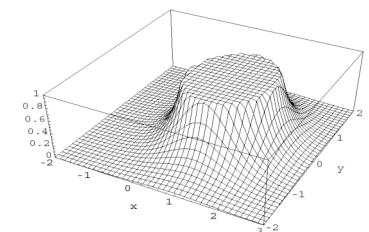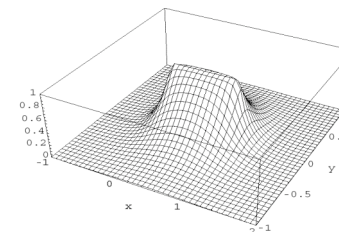$$P_{R_t} = \exp\{k[d(S,D) - d(R_t,D) - t]\}$$

Performance results
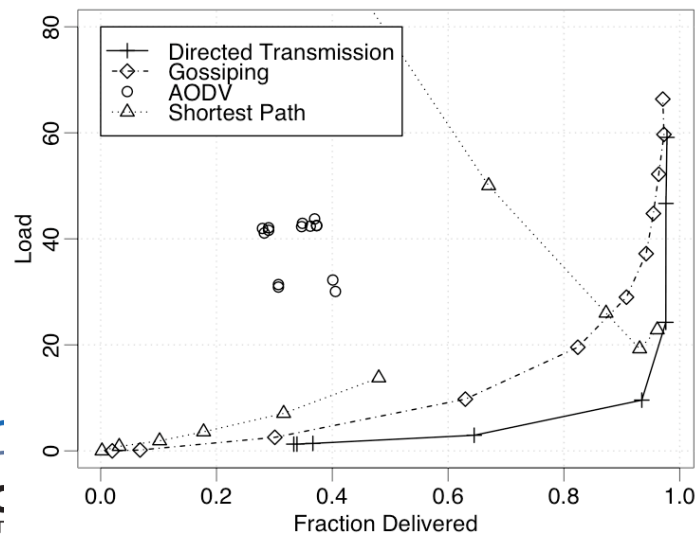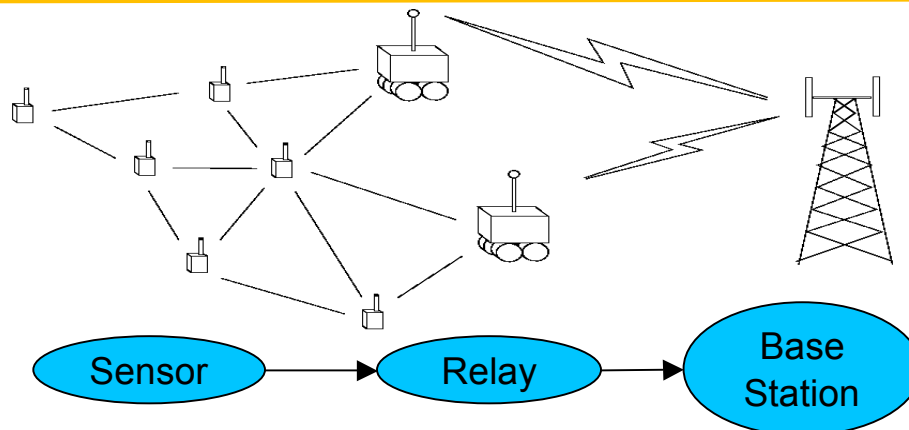

General routing protocol classification

Load Profiles

Dir. Transmission   Dest. Attractor

Hierarchical sensor networks: Maneuverable relays

[S. Eidenbenz, L. Kroc, J.P. Smith, PerCom Workshop PeRSens 2005]

Sensor → Relay → Base Station

Objective: Given sensor positions, find relay positions such that the energy consumption E_S is minimized under shortest path routing

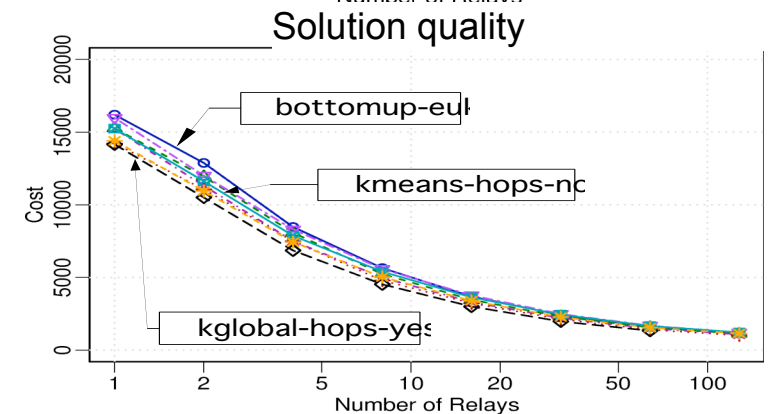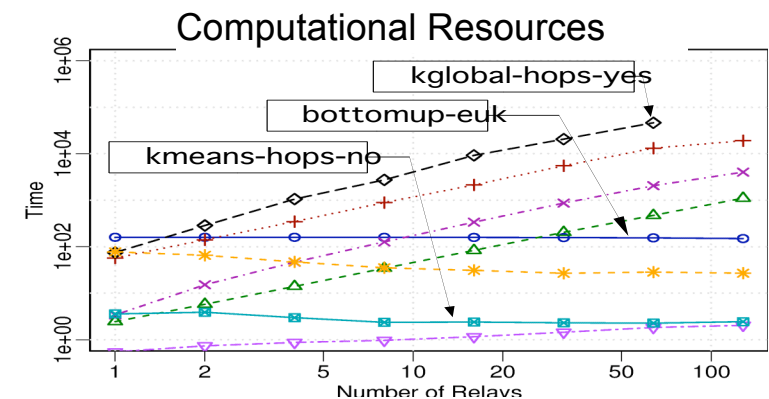$$E_S = \sum_{i=1}^{n} f_i \cdot d_i \cdot E_U$$

f = # packets generated
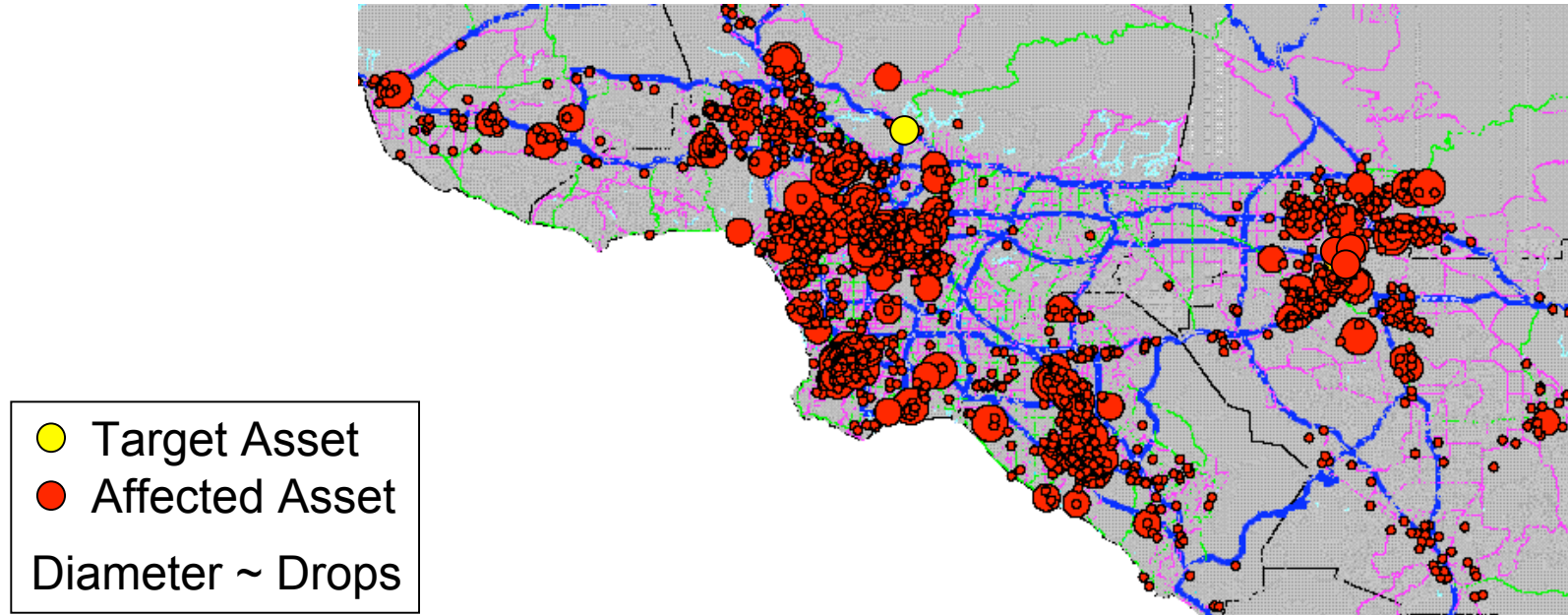d = hop count to relay
E_U = energy per transmission
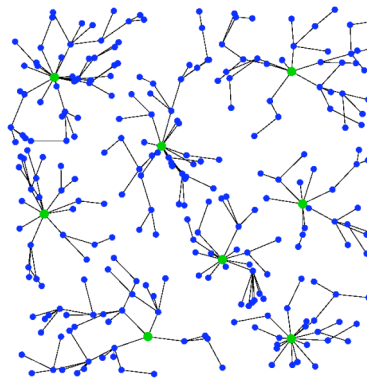
Results
(1) Problem is NP-hard
(2) Trade-off quality vs. computational resources of various heuristic algorithms

Computational Resources

kglobal-hops-yes
bottomup-euk
kmeans-hops-no

Solution quality

bottomup-euk
kmeans-hops-no
kglobal-hops-yes

Communications Asset Criticality Ranking (fictitious example)

Legend:
- ○ (yellow) Target Asset
- ● (red) Affected Asset

Diameter ~ Drops

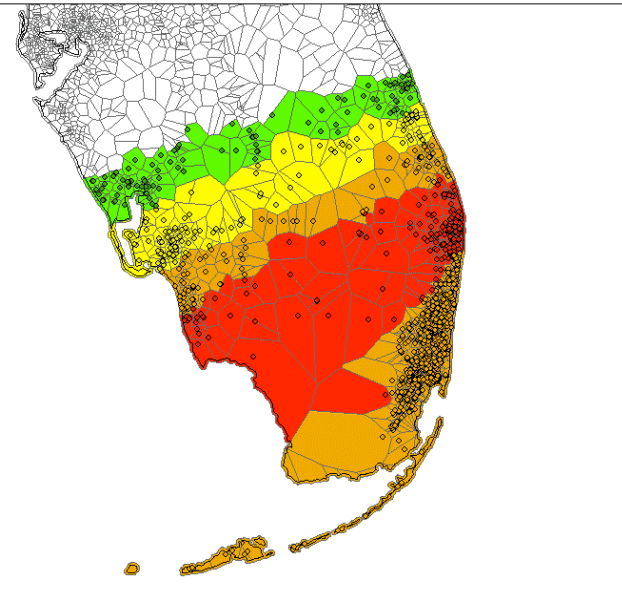Operator: California Bell
Address: 9999 California Rd,
California, CA 99999
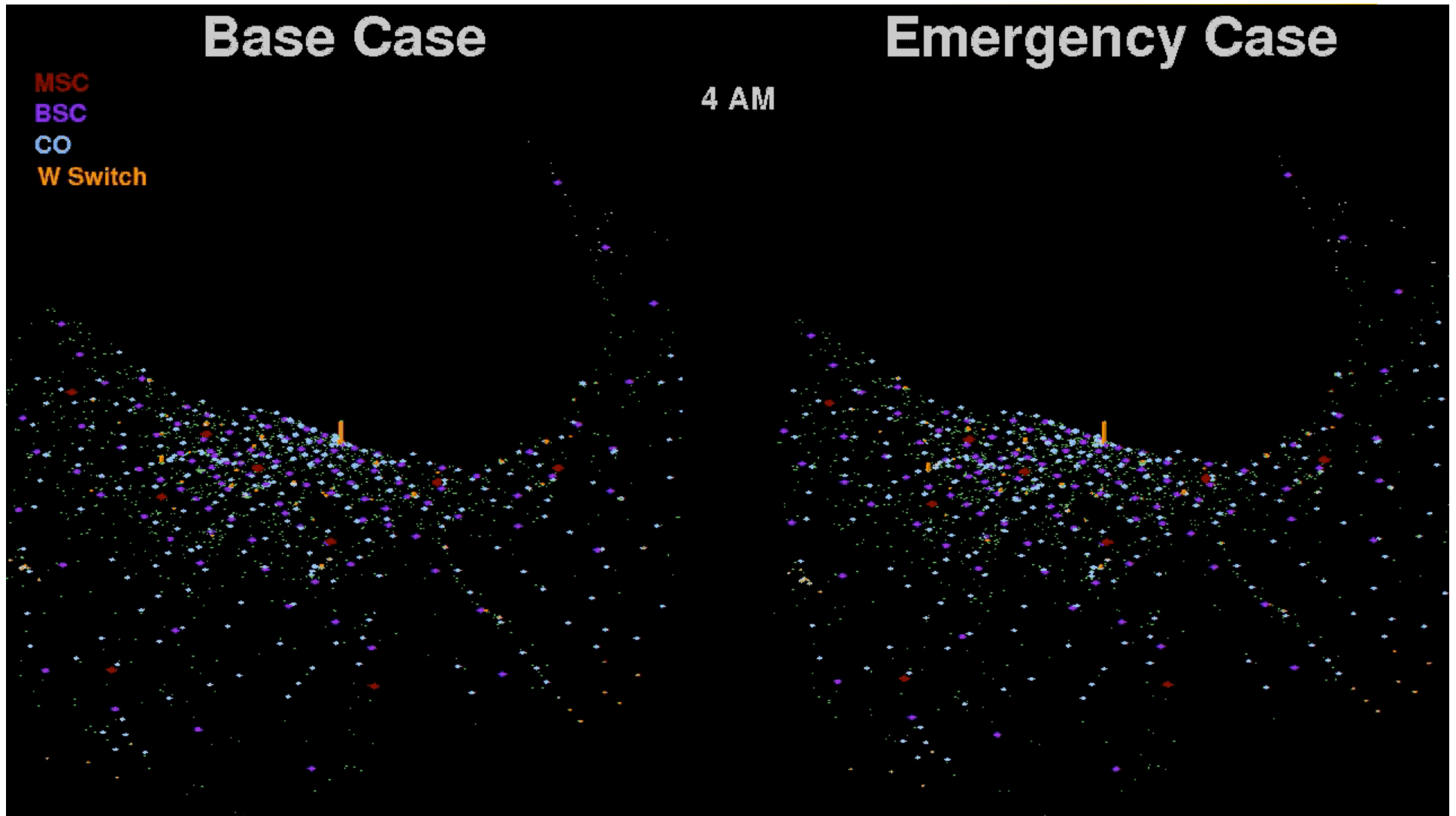2,234,234 total lost calls
$3,622,324 business loss

Los Alamos
NATIONAL LABORATORY
EST.1943

NNSA

Hurricane Asset Damage Analysis

Rita Damage to Cell Phone Towers
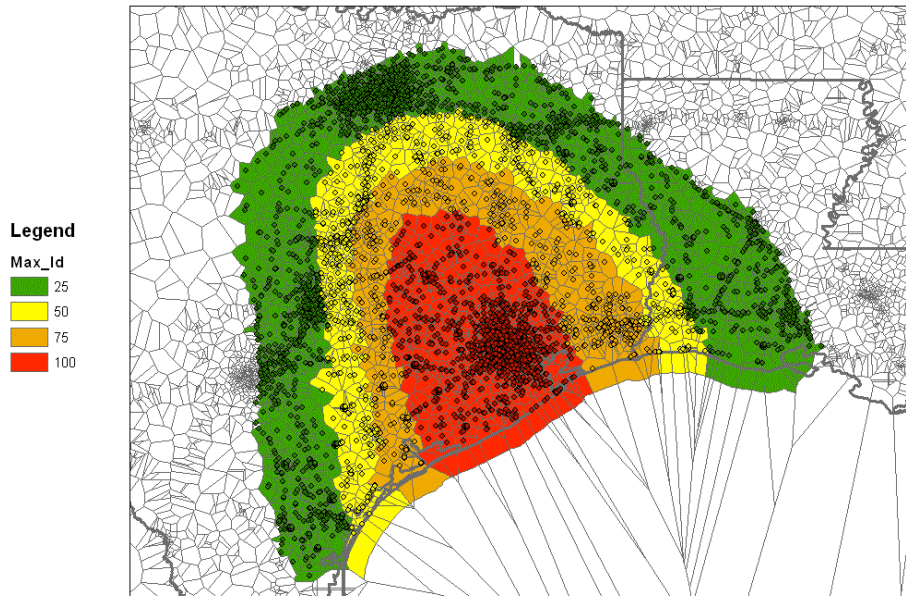
Wilma Damage to Cell Phone Towers

Legend

Max_Id

- 25
- 50
- 75
- 100

PSTN Example Scenario – Network Loads

Generic Methods

Anders Hansson

and

Hristo Djidjev, Stephan Eidenbenz, Nick Hengartner,
Gabriel Istrate, Shiva Kasiviswanathan,
Matthew Nassr, Gopal Pandurangan, Rajiv Raman,
Christian Reidys, Zoya Svitkina, Dave Tallman,
Sunil Thulasidasan

Goals of Project

Systematic development of scalable disaggregate models of functioning infrastructures

1. Quick turnaround; Severe time constraints; Scalability

2. Conclusions transferability; General guidelines (normative properties or policies) applicable across regions are desired

3. Data transferability; Detailed data unavailable; Generic information for specific localities must be generated

Overview of Accomplishments (1/3)

Analysis and modeling of large-scale networks, "modern graph theory," is central to Generic Methods

> Fundamental algorithms for network analysis
  - Extensive software library
  - Betweenness • Diameter • Expansion ratio • Overlap ratio
    • Shattering coefficient • Capacity under linear constraints
    • Distance-2 matching • Distributed local spanning trees
    • Fast sampling based algorithms for estimating clustering coefficients
    • Shortest path distribution • Dominating sets

> 20–30 refereed publications
  - Including: SODA • Nature • Annals of Combinatorics
    • SIGMETRICS • DIMACS • Discrete Mathematics • etc.

Los Alamos
NATIONAL LABORATORY
EST.1943

NNSA

Overview of Accomplishments (2/3)

Structural analysis of massive social networks
- New fundamental measures for understanding epidemics

Combinatorial models for epidemiology policy planning
- Novel formulations and effective algorithms for quarantining, vaccination, and sensor placement

Random graph models of real social networks
- Model validated w.r.t. key measures
- Efficient data structure
- Very fast generative algorithms

Overview of Accomplishments (3/3)

Analysis of urban IP networks; City-nets

- Improved techniques for network inference
- Validated structural model of city-nets
- Exposed the vulnerability of city-nets; Principled approach to asset ranking
- Efficient algorithm for detecting denial-of-service attacks

Network flow interdiction analysis

- Fast algorithm for identifying critical links/nodes

Evacuation planning (SWATH & TOPOFF IV)

- Fast methodology for estimating evacuation times

Structural Analysis of Massive Social Networks

High expansion ratio: Contagious diseases spread very fast

- Early detection is imperative
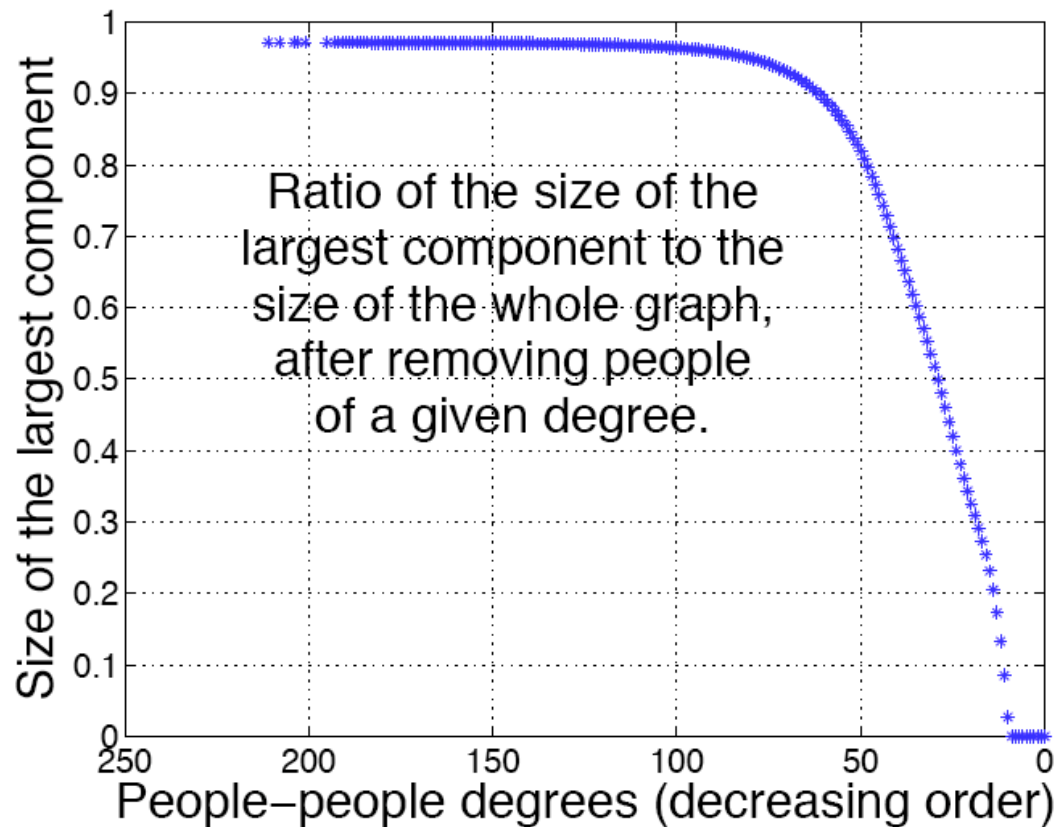- Time of withdrawal to home most important factor

$$\frac{\left| arcs\,(u,v) : u \in S, v \notin S \right|}{\left| S \right|}$$

People-people graph
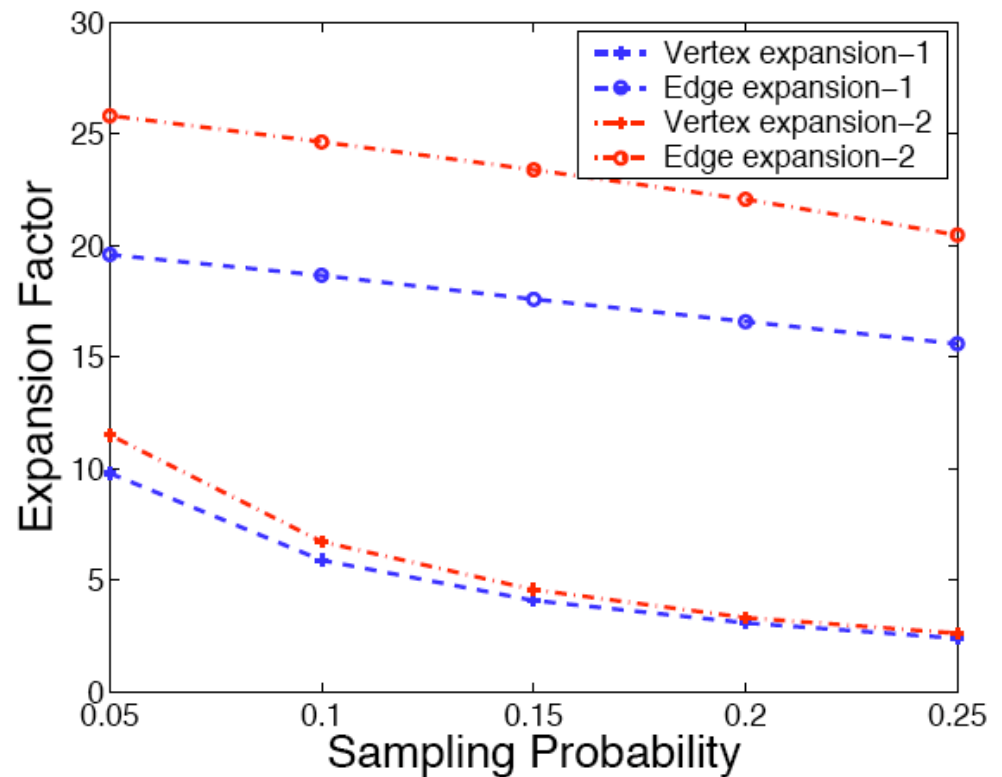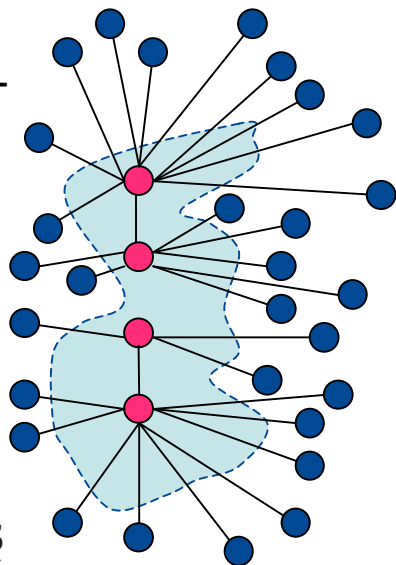
Structural Analysis of Massive Social Networks

Hard to shatter: Vaccinating all people of degree greater than some threshold not effective

Ratio of the size of the largest component to the size of the whole graph, after removing people of a given degree.

Random Graph Models of Real Social Networks

Significant reduction in response time

- Current approach: several days
  - Data preparation: at least a few days
  - Synthetic population generation: a few hours
  - Activity generation: LA took 14 hours (30-CPU cluster)

- Generic methods: a few minutes

Random Graph Models of Real Social Networks

Cultural and structural similarities between cities
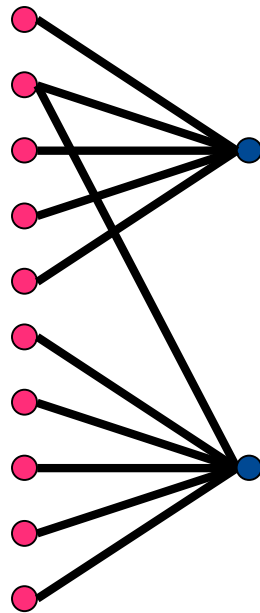
– For example: Local properties, such as degree distributions
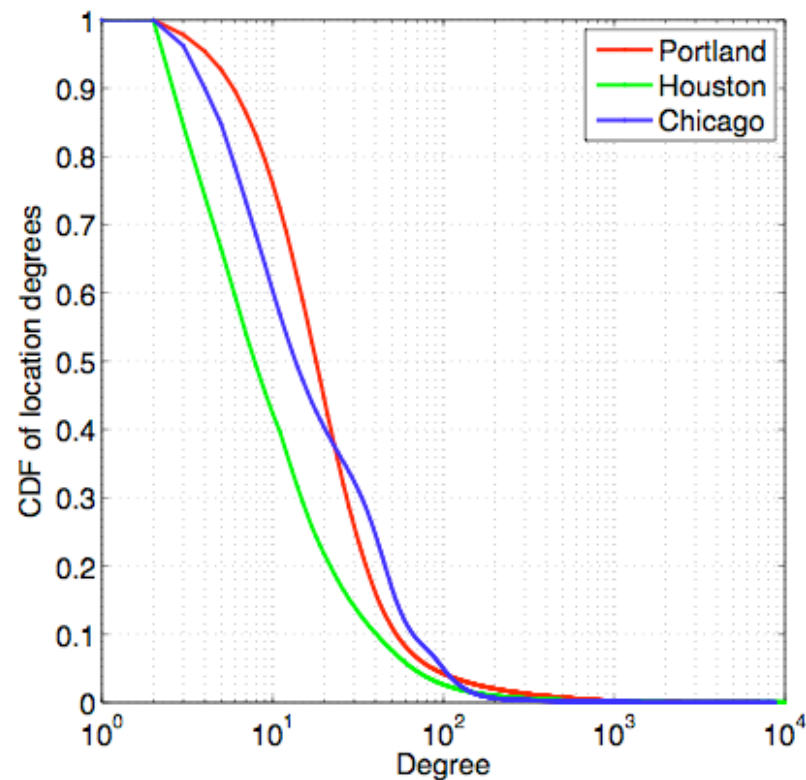
Random Graph Models of Real Social Networks

Parametric characterization of cities

"Tell me the number of locations/people, and I can generate people's contact patterns"
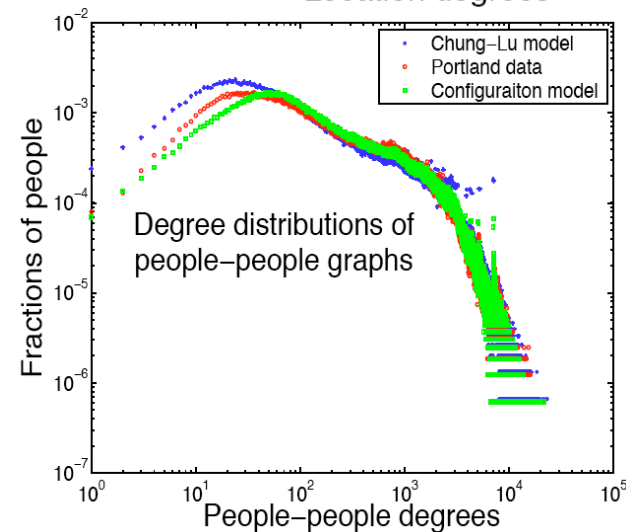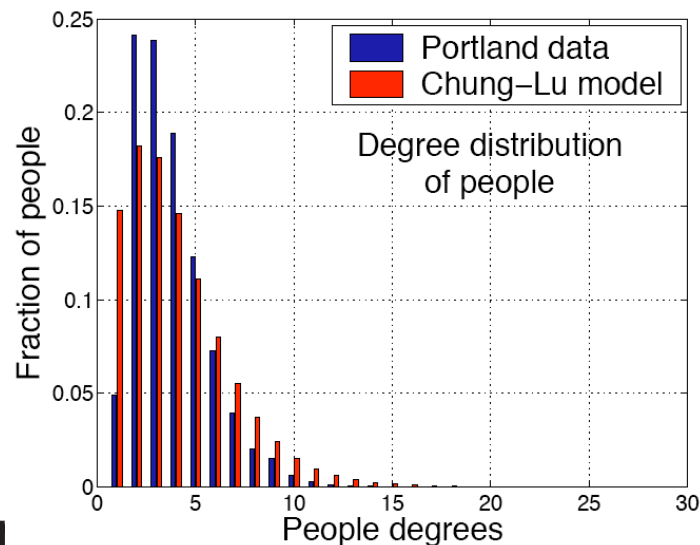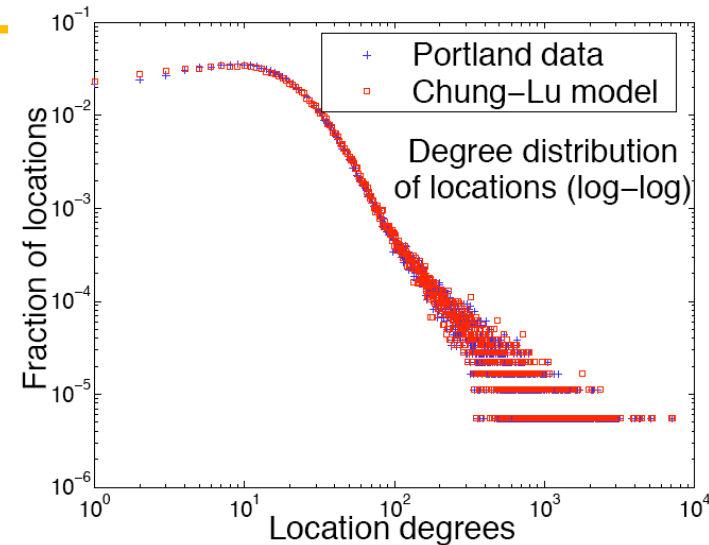
People-location (PL) graph

The Aiello-Chung-Lu (ACL) model creates an instance of the PL graph

Random Graph Models of Real Social Networks

Fast generative algorithms

- – Naïve ACL takes 10 hours
- – Our ACL takes 1 minute
- – High agreement with real data

Degree distribution of locations (log–log)

Degree distribution of people
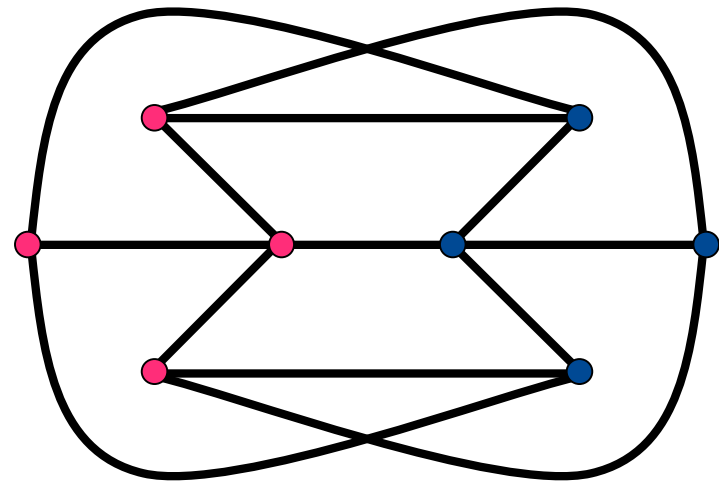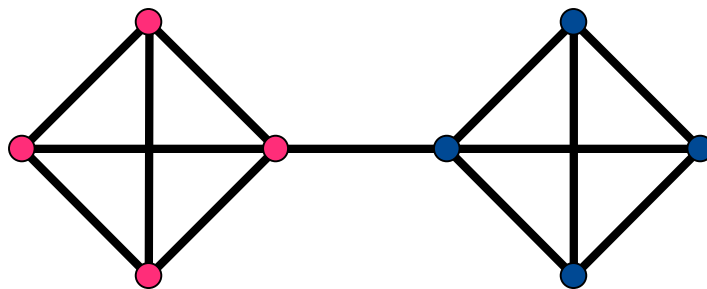

Degree distributions of people–people graphs

Random Graph Models
# of Real Social Networks

Degree sequences reflect local properties

On-going work: Extend ACL to match global properties, such as
community structure

– Algorithms: greedy, shingles, gravity, min-cut, spectral

Analysis of City-nets

- Executed 5 million trace-routes

- Analyzed 11 of the 25 largest cities in the U.S.

- A few routers (10–40) carry over 80% of the traffic

- If 0.3% of the routers are affected by failure/attacks traffic reduces by 80%

- City-nets are **much more vulnerable than the global Internet**
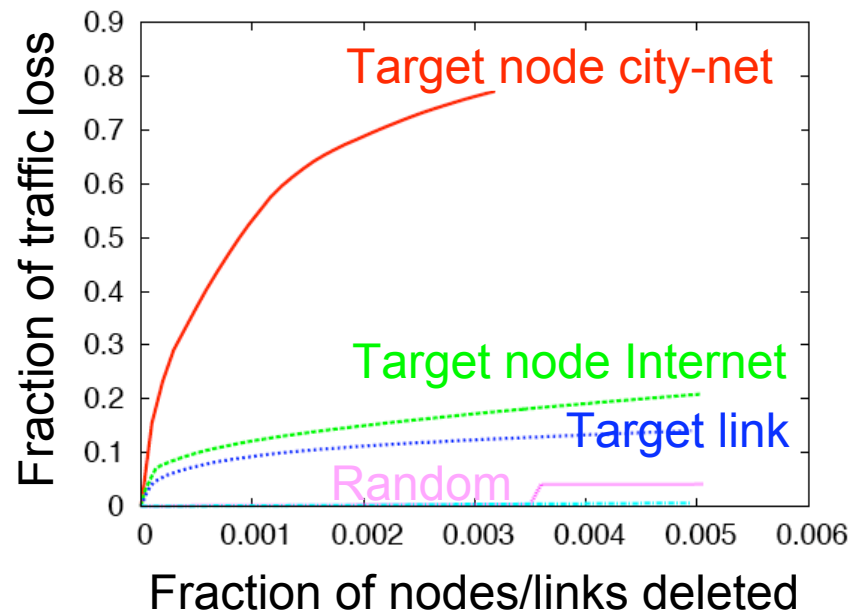
Analysis of City-nets

- Confluent paths: not diverging after converging once
  - Almost all city-net paths are confluent

- Developed efficient monitoring capabilities
  - A small set of paths is guaranteed to be affected by any large-scale denial-of-service attack—independent of the network size